

HPE Cray Supercomputing EX254n

8x GH200

Grace Hopper

blade

OVERVIEW

—	0.8 TB	—	—
Peak Dense	Aggregate Memory	Aggregate Bandwidth	Aggregate Power

CONFIGURATION

Accelerator	GH200 96GB HBM3
Count	8
Memory per GPU	96 GB
Bandwidth per GPU	4.0 TB/s

POWER & EFFICIENCY

Aggregate Power	—
TDP per GPU	1000 W
Efficiency	—
Aggregate Bandwidth	—

SYSTEM DETAILS

Vendor: HPE	Family: Cray Supercomputing EX	Architecture: Grace Hopper
Form Factor: blade	Release Year: —	Process Node: 4nm

INTERCONNECT

HPE Slingshot, with four injection ports per node and up to eight 200Gbit/sec ports per blade, which works out at one NIC per GPU; the fabric is a dragonfly topology built from 64-port Slingshot switch blades rated at 12.8 Tb/s apiece

NOTES

HPE's Grace Hopper accelerator blade for the Cray Supercomputing EX, and a genuinely different shape from the eight-GPU servers that fill the rest of this catalogue. A single-slot blade holds two four-socket GH200 nodes, so eight Grace Hopper superchips share one slot of a chassis that takes eight blades, and a full EX4000 cabinet takes 64 blades across 8 chassis. Memory is 768 GB of HBM3 across the eight GPUs at 96 GB each, alongside 960 GB of LPDDR on the Grace side at 120 GB per CPU. Networking is HPE Slingshot rather than InfiniBand or Ethernet: four injection ports per node and up to eight 200Gbit/sec ports per blade, which HPE notes is configured as one NIC per GPU by fitting four switch blades per chassis. The whole blade is cooled by cold plate with no fans anywhere in the cabinet, which is how the EX4000 stays sealed and returns no heat to the room. Storage is minimal by design, one optional NVMe M.2 per node. HPE publishes no power figure for the blade and no throughput figure for any blade in this document; the cabinet-level ratings it does give are facility input in kVA for an enclosure that may hold any mix of blades.